

Fordítások statisztikai alapú minőségvizsgálata tartalomelemzéssel

Puskás László

PTE BTK, Pszichológia Doktori Iskola
laszlopuskas@gmail.com

Kivonat: A tanulmány egy olyan eljárást mutat be, amely egy olasz szöveg magyar fordításának statisztikai jellemzőit vizsgálva igyekszik olyan általános statisztikai összefüggések bemutatására, amellyel meghatározott típusú szövegekben, a statisztikai jellemzők alapján, kiszűrhetők a hibás fordítást tartalmazó szövegrészek, illetve bizonyos típusú hibák. Az eljárás során az olasz szöveget meghatározott méretű szövegrészekre bontjuk, amelyeket azok magyar fordításával vetünk össze, a szövegrészek statisztikája alapján. Azt feltételezzük, hogy az eljárás az olaszon kívül más nyelvre is alkalmazható, azzal a megkötéssel, hogy a vizsgált statisztikai paraméterek eltéréseinek általános szabályait az adott nyelvre is ki kell dolgozni.

1 Bevezetés

A tanulmány egy az önéletrajzi emlékezet körébe tartozó olasz művet, és annak magyar fordítását megvizsgálva igyekszik meghatározni, hogyan szűrhetők ki bizonyos fordítási hibákat tartalmazó szövegrészek, statisztikai alapon. A cikket egyben vitaindítónak is szánom. Tanulmányomban a következő feltevések igazolására törekszem:

I. meghatározott típusú szövegek esetén, az olasz szövegrészekben szereplő szavak száma szinte mindig nagyobb a magyar szövegrészekben szereplő szavak számánál;

II. az olasz és a magyar szövegrészben szereplő szavak számának eltérése arányában általában jól behatárolható, de a mondat szintjén nem, csak a szövegrész szintjén alkalmazható;

III. az együttjárások elsősorban a történetek elbeszéléséhez, vagyis a narratív szemléletmódhoz köthetők;

IV. a szavak számának eltérése a szövegben szereplő szófajok arányainak eltéréseivel is együttjár;

V. az eljárás általános alkalmazása lehetővé teszi, hogy olasz és más idegen nyelvű szövegek hibás fordítását nagy valószínűséggel felismerjük, azaz a módszerrel nyelvfüggetlenül hasonlítsunk össze idegen nyelvű szövegeket magyar nyelvű fordításaikkal.

2 Nyelv és kultúra kapcsolata, nyelvi relativitás

A nyelvészetben már korábban is foglalkoztak azzal a gondolattal, miszerint a különböző nyelvek különböző gondolkodásmódokat takarnak, és meghatározzák használatuk világképét. Ezt a gondolatot először Wilhelm von Humboldt, német nyelvész és polihisztor, vetette fel a XIX. század elején. Később, a XX. század második felében, a kulturális antropológián belül indult meg a nyelvek kulturális összehasonlító vizsgálata, mely Edward Sapir és Benjamin Whorf nevéhez fűződik. A Sapir-Whorf hipotézis szerint a nyelv struktúrája és szemléletmódja meghatározza a valóságlátást és a külvilágból jövő ingerek érzékelését. Előadásomban ennek a gondolatnak egy sajátos megközelítésével kívánok foglalkozni: hogyan adható át egy gondolat két különböző szerkezetű nyelv között anélkül, hogy az átadott gondolat megváltozna, és statisztikai módszerekkel hogyan szűrhetők ki a fordítási hibák. A magyar nyelv a legtöbb európai nyelvtől különbözik. Az eltérő szerkezetű nyelvek fordítása során egy eltérő szerkezetű szöveg jön létre. Mivel az eltérések általában szisztematikusak, így statisztikai alapon vizsgálhatók. Feltételezésem szerint a nem megfelelő módon, szerkezetben átadott fordítás a megfelelő szerkezetű fordítástól eltérő statisztikai paraméterekkel rendelkezik, amely számszerűsíthető, ezzel kimutatva a hibás fordítást.

A Sapir-Whorf hipotézis szerint a világot a rendelkezésünkre álló fogalmakból tudjuk megérteni, és mivel ezeket a nyelv biztosítja számunkra, a más nyelven beszélők másképp látják a világot, más „megismerési univerzumban” élnek. Az elméletnek egy szélesebb körben elfogadott, enyhébb változata szerint a nyelvi különbségek hajlamossá tehetik az embereket, hogy másképp lássák a világot. [7] A hipotézist sok bírálat érte, miszerint a nyelv kultúra meghatározó szerepe nem igazolható, ugyanakkor az kevésbé vitatható, hogy a más nyelvet beszélők között kulturális különbségek vannak. A különböző nyelvek nemcsak eltérő szókinccsel és fogalmi repertoárral rendelkeznek, hanem az eltérő nyelvi szerkezettel a gondolatok szerveződésének egy egészen más módja valósul meg. Azt feltételezem, hogy az eltérő nyelvi szerveződések olyan rendszert alkotnak, amelyekben az eltérő nyelvi szerkezetek egyrészt megfeleltethetők egymásnak két különböző nyelv között, másrészt a szisztematikus eltérések részben számokkal mérhetővé tehetők.

A nyelvek közötti különbségnek azonban van egy másik vetülete is: nemcsak eltérő nyelvi szerkezetről van szó, hanem a gondolkodásnak egy olyan sajátos formájáról, amely ugyanarról a kérdésről, akár egészen más felfogásban fejeződik ki különböző nyelveken. A Sapir-Whorf hipotézis igazolására több kísérletet is végeztek. Ezek közül az egyikben az Egyesült Államokban élő kétnyelvű japán nőkkel készítettek interjúkat, akik mind a két nyelvet egyaránt jól beszélték. Két, egymástól elkülönült interjút készítettek, az első alkalom japán nyelven folyt, míg a második angol nyelven. Az interjú során azt a feladatot kapták az interjúalanyok, hogy egészítsék ki ugyanazokat a mondatokat, első alkalommal japánul, másodjára angolul. A gondolkodásmódban való eltérést, amikor különböző nyelven kellett megoldani a feladatot, a következő példák szemléltetik:

Amikor a vágyaim összeütközésbe kerülnek a családom vágyaival...

...ez nekem boldogtalanságot okoz. (japán)

...megpróbálom valóra váltani a vágyaimat. (angol)

Az igazi barátoknak...

...segíteniük kell egymást. (japán)

...őszintének kell lenniük egymáshoz. (angol) [6]

2 A vizsgálati módszer

A tanulmányban vizsgált szövegrészek Fabrizio Ciano Amikor nagyapa agyonlövötte apát című könyvéből valók. [4], [5] Egészen pontosan annak első hat fejezete, amelyet meghatározott módszer szerint ötvenhárom szövegrészre osztottam, amely megfelelő esetszámnak bizonyul, ahhoz, hogy ezeken a szövegrészeken vizsgálatokat folytatva statisztikailag is értékelhető eredményeket kapjunk.

Az eljárás módszere: a szöveget meghatározott méretű szövegrészekre bontjuk, majd az adott szövegrészt és annak fordítását statisztikai szempontból összehasonlítjuk. A szövegrészek hosszúságának megállapításánál két szempontot kellett figyelembe venni. A kiválasztott szövegrész ne legyen túl hosszú, mert ebben az esetben az esetleges hibák, a szavak nagy száma miatt, elveszhetnek, minthogy az apróbb hibák magas szószám esetén nem befolyásolják lényegesen a szövegrész statisztikáját. Másrészt túl rövid szövegrész esetén a két összehasonlított szövegrészben szereplő szavak aránya nagyobb ingadozást mutathat, és kevésbé közelít az átlaghoz. A vizsgált mű esetében általában oldalanként haladtam. Ettől két esetben tértem el. Ha egy oldalon kevesebb, mint húsz sor volt, akkor az előtte lévő oldallal közösen vizsgáltam. Az egyik oldalról a másikra átnyúló mondatokat minden esetben azzal az oldallal együtt vizsgáltam, ahol elkezdődtek. A fejezetcímeket az összehasonlító vizsgálatból kihagytam, mivel magára a szövegre voltam kíváncsi. Természetesen a könyvben szereplő oldalakat, mint a vizsgálat tárgyát képező szövegrészeket, nem lehet állandó statisztikai egységnek tekinteni, még akkor sem, ha ebben az esetben beváltak. Megvizsgálva a kiválasztott szövegrészeket, nagyrészt 150 és 400 szó közötti egységeket alkotnak, átlagos hosszúságuk 258,1 szó. Úgy tűnik, hogy nagyjából a megadott méretű szövegrészek mellett, az eljárás nagy pontossággal lefolytatható, és statisztikailag értékelhető adatokat eredményez.

A statisztikai adatokat Word program segítségével kérdeztem le. A szavak számán kívül rögzítettem a szövegrészekben szereplő karakterek számát is, mind az eredeti, mind pedig a lefordított szövegrészekben. Az eljárás lefolytatásához olyan segédprogram is előállítható, amely részekre bontja az eredeti szöveget, viszont ebben az esetben is szükséges a szövegrészek fordításának kijelölése, megfeleltetése, illetve ellenőrzése. Ilyenkor a meghatározott hosszúságú, vagy ahhoz közelítő szövegrészek kiválasztását célszerű bekezdéstől bekezdésig kijelölni.

A karakterek számának elérése az olasz és a magyar szövegrészekben rendszertelen ingadozást mutatott, így ezt az eljárás során nem használtam fel. Annyit mindenestre érdemes megjegyezni, hogy bár az olasz szövegrészek szószáma minden esetben nagyobb volt a magyar szövegrészek szószámánál, a vizsgált teljes szövegben a karakterek száma szóköz nélkül a magyar szövegben 2,5 százalékkal magasabb volt, mint az eredeti olaszban. A karakterek száma szóközőkkel együtt pedig egy elhanyagolható 0,2 százalékos eltérést mutatott a magyar szöveg javára, tehát az olasz nyelvű

szöveg ugyanazt a gondolatot gyakorlatilag ugyanolyan hosszan fejezte ki írásban, mint a magyar.

Az idegen nyelvű szövegrészek és magyar fordításuk összehasonlításánál meg kellett határoznom, hogy a vizsgálat szempontjából mit tekintek önálló szavaknak. Erre azért volt szükség, mert az olasz nyelv, meghatározott szabályok szerint, gyakran rövidít aposztróffal bizonyos szavakat, elsősorban a névelőket és bizonyos esetekben a tárgyas névmásokat. Mivel az aposztróf olyan írásjel, amit, a többitől eltérően, nem szököz különít el az írásjelet követő szótól így ezekben az esetekben a Word program statisztikai szempontból a két szót egynek tekinti. Így létre kellett hoznom egy olyan állományt, amelyben a kiválasztott szövegrészekben az aposztróf és az azt követő betű közé szöközt tettem. Ezt követően megvizsgáltam az így módosított szövegrészek jellemzőit. A módosított olasz szövegrészekben és azok magyar fordításában megszámoltam a program segítségével a szavakat, és megvizsgáltam azt, hogy az olasz szavak száma milyen arányban tér el a magyar szövegészben szereplő szavak számától. Statisztikailag megvizsgáltam azonban azt is, hogy a javítatlan szövegrészek szempontjából, ahol az aposztróf előtti és utáni szót a program egyként kezeli, hogyan alakul ez az arány. (Míg a módosított szövegrészekben az átlagos szószám 258,1 szó volt, ez a módosítás előtt, átlagosan 251,2 szó volt.)

A megvizsgált szövegrészekben az olasz szavak száma minden esetben nagyobb volt az azok magyar fordításában szereplő szavak számánál. Az eltérés 8,4 és 26,7 százalék között mozgott, tehát egy jól behatárolható 18,3 százalékos intervallumban. Ha mind a hat kiválasztott fejezetet osztatlanul, egyben vizsgáljuk, akkor az olasz szöveg szavainak száma 17,2 százalékkal haladta meg a magyar fordítás szószámát. Ha a javítatlan szövegrészekkel hasonlítom össze magyar megfelelőjüket, akkor az olasz szövegben szereplő szavak száma legalább hat és legfeljebb 24,8 százalékkal haladta meg a magyar szövegrészek szószámát, hozzáátve azonban azt, hogy ebben az esetben nem nyelvtani ételemben használok a szó kifejezést, hanem statisztikai egységként, két különálló szöveg összehasonlítására. Ebben az esetben is egy a korábbihoz hasonló intervallumba esnek a szavak, amely 18,8 százalékos értéket mutat.

A fejezeteket és azok fordításait külön is megvizsgálva a következő eléréseket találtam a szavak számában az olasz szöveg javára:

- 1. fejezet: 15,9 százalék;
- 2. fejezet: 14,8 százalék;
- 3. fejezet: 15,7 százalék;
- 4. fejezet: 15,7 százalék;
- 5. fejezet: 19,2 százalék;
- 6. fejezet: 19,8 százalék.

A fejezetek és fordításuk szószámának elérése egy jól behatárolt ötszázalékos intervallumban mozog. A fejezeteket hét-tizenhárom szövegrészre osztottam. Ha a szövegrészek statisztikai adatai alapján szeretnénk következtetéseket levonni a fordítás helyességére, arra van szükségünk, hogy a százalékos eltérések viszonylag szűk intervallumba essenek, hiszen minél szűkebbre szabott ez az intervallum annál érzékenyebben reagál az eljárás a szavak számában való eltérésre. A fejezetek megfelelnek ugyan ennek a kitételnek, a hosszúságuk miatt azonban nem válnának be érzékeny indikátorként, viszont egy összetett vizsgálati eljárás egyik összetevőjeként, más feltételek együttes megléte estén jól használhatók. A szövegrészek 18,3 százalékos intervalluma is elég jól behatárolt, de kerestem annak a lehetőségét, hogyan tudnám

csökkenteni, így megvizsgáltam a szavak számaránybeli eltérésének szélsőértékeit, mégpedig az alsó és felső tizedbe tartozó értékeket. Ha az alsó és felső tizedbe tartozó értékek nélkül nézzük a szavak számbeli elérésének arányait, akkor egy tizenkettő és huszonöt százalék közé eső, tizenhárom százalékos intervallumot kapunk, ebbe esik bele a megvizsgált szövegrészek nyolcvan százaléka.

Ha tehát hosszabb szövegrészeket vizsgálunk, akkor a fordítások ellenőrzésére, minőségvizsgálatára a következő eljárást célszerű alkalmaznunk: a szöveget először nagyobb gondolati egységekre bontjuk, amelyeket azután a korábban ismertetett módon, meghatározott méretű szövegrészekre bontunk. A szövegrészek vizsgálatánál a tizenkettő és huszonöt százalék közötti elérést érdemes minden esetben elfogadnunk, a száznyolc és száztizenkettő százalék közötti eltérést, valamint a huszonöt és huszonhét százalék közötti eltérést pedig abban az esetben, ha a nagyobb gondolati egység és annak fordítása között a szavak számában való eltérés tizennégy és huszonegy százalék közé esik (ebben az esetben a vizsgálatnál kapott szélsőértékeket, fölfele és lefele egy-egy százalékkal kitoltam). Így egy statisztikailag jól körülhatárolt, könnyen kezelhető és érzékeny indikátort kapunk. Ha valamelyik szövegrész nem felel meg a meghatározott statisztikai paramétereknek, szükséges a fordítás ellenőrzése!

Amikor az eltérő hosszúságú fejezeteket kezdtem vizsgálni, akkor ezt abból a megfontolásból tettem, hogy ezek a részek önálló narratív egységeket alkotnak, így önálló narratív struktúrával rendelkeznek, és ezek a struktúrák, azt feltételeztem, hasonlóságuk révén jóval kisebb statisztikai eltéréseket fognak mutatni egymáshoz képest, mint az őket alkotó szövegrészek. A fordítások statisztikai alapú összehasonlíthatóságát pedig részben magának a narratív struktúrának tulajdonítottam.

Bruner az emberi értelem működésének két módját különbözteti meg: az egyik a logikai-tudományos vagy más néven paradigmatis mód, a másik pedig a narratív mód. A paradigmatis gondolkodásmódról az évezredek folyamán tekintélyes tudás halmozódott fel, és tudományos egzaktuság uralja, míg a narratívum ezzel szemben „az emberi szándékok viszontagságaival foglalkozik”. A történeteknek látszólag végtelen lehetséges elbeszélési módja van, azonban ez még sincs egészen így. „Egyes nézetek szerint az életszerű narratívum egyfajta kanonikus vagy „legitim” szilárd állapottal kezdődik, amely törést szenved és válságba kerül, hogy azután orvoslást nyerjen, s e ciklus megismétlésére nyitva a lehetőség.” [3] Azt feltételezem, hogy nemcsak a narratívum, hanem az azt tartalmazó szöveg formai sajátosságai is hordoznak egy olyan struktúrát, amely a narratívumot hordozó nyelv szerkezetével és statisztikai jellemzőivel párhuzamosan vizsgálható, és felhasználható a más nyelvre fordított megnyilatkozás szerkezetével és statisztikai jellemzőivel való összevetésre.

Miért kötöm inkább a narratív gondolkodásmóddhoz a szóstatisztikák összevethetőségét? Könnyű belátni ennek az elképzelésnek a valószínűségét, ha belegondolunk abba, hogy a képletekkel, anyagnevek felsorolásával, kísérletek logikai leírásával foglalkozó logikai-tudományos gondolkodásmód nem olyan szerkezettel rendelkezik, amely lehetővé tenné a vizsgált eljárás tökéletes alkalmazását (az anyagnevek és képletek a fordításkor nem eredményeznek mérhető változást a szavak számában, és a leírások igeidő használta kisebb változatosságot mutathat). Ha viszont egy önálló szöveg narratív struktúrával rendelkezik, annak szerkezete, más nyelvre való lefordításakor szisztematikusan változik, és ez a változás statisztikailag mérhető eredményt hoz magával.

Ha nem nagyobb gondolati egységeket vizsgálunk, vagy nem kívánjuk nagyobb gondolati egységekre bontani a vizsgálandó szöveget, akkor a szövegrészek vizsgálatakor egy árnyaltabb kategorizálást használhatunk, a tizenkettő és huszonöt százalék közötti különbségre azt mondhatjuk, hogy a megadott határértékek között van, míg a tanulmányban az alsó és felső tizedbe eső eredményeket a statisztikailag elfogadható kategóriába soroljuk. A nyolc százalék alatti és a huszonhét százalék feletti tartományba eső eltéréseknél minden esetben javasolt a fordítás ellenőrzése.

Kizárható-e a fordítás helyessége, ha a tanulmányban megadott határértékeken kívülre esik a szavak eltérési aránya? Természetesen a fordítás helyessége ebben az esetben sem zárható ki teljesen, még akkor sem, ha ez statisztikailag nem valószínűsíthető, ezért is fogalmaztam úgy a bevezető részben, hogy az olasz szöveg szószáma „szinte” minden esetben nagyobb a magyar szöveg szószámánál. A határértékeken kívül eső szövegrészek ellenőrzése azonban minden esetben célszerű. Az alkalmazott eljárás nem arra szolgál, hogy egy adott szövegrészről megállapítsuk, hogy annak fordítása helyes-e, sokkal inkább arra, hogy kiszűrjük vele az egyértelműen hibásnak vélelmezhető szövegrészeket, és az esetleges fordítási hibákat a szöveg ellenőrzését követően kijavítsuk.

Milyen típusú hibák kezelésére alkalmas az eljárás, és melyekre nem? A szavak félrefordításából eredő hibákat nem tudjuk kiszűrni ezzel az eljárással, hiszen a statisztikában nem jelenik meg semmilyen eltérés. Az összehasonlított szövegrészek szószámának eltérése részben a magyartól eltérő szerkezetekből, részben pedig a kifejezések, szófordulatok eltéréseiből adódik, amely sok esetben szintén az eltérő nyelvet használók szemléletbeli különbségeire vezethető vissza. A szövegrész nem megfelelő szerkezetben történő átadása nem adja vissza a szavak arányainak eltérését, így a félrefordított szöveg könnyen kiszűrhető. Egy másik hibalehetőség, amikor a fordításból kimarad valami. Egy mondat, vagy akár egy fél mondat kimaradása is olyan elérést eredményezhet az eredeti arányokhoz képest, amely könnyen kiszűrhetővé teszi a hibát.

3 A nyelvhasználat és az eltérő nyelvi szerkezetek hatása az eredeti és a lefordított szöveg szóstatisztikája közötti különbségre

Nem csak két különböző nyelven megnyilatkozó ember beszéde, írása között találunk különbséget, hanem gyakran az azonos nyelvet beszélők megnyilatkozásaink formája is eltérhet egymástól. Az eltérő nyelvhasználatot nem csupán az eltérő gondolkodásmód okozza, hanem az adott társadalmon belüli rétegződésbeli, tanultságbeli és szocializációs különbségek. Ezen kívül még számolnunk kell a különböző nyelvjárásokban beszélők nyelvi normájának különbözőségével is, amely bizonyos nyelvek esetében markáns különbségeket mutathat. A felsorolt különbségek részben kulturális jellegűek, de bizonyos esetben az adott nyelv tökéletlen elsajátítása is eredményezheti. Meg kell tehát vizsgálnunk, hogy a felvázolt elmélet alkalmazható-e ezekben az esetekben is, és ha igen, akkor milyen megkötésekkel, valamint azt is, hogy az adott nyelven belüli különbségek miben térnek el az adott nyelv sztenderdjétől.

A különböző társadalmi rétegek eltérő nyelvhasználatának társadalmi hátrányokat továbbörökítő hatásával már több szerző is foglalkozott. Basil Bernstein ennek első-

sorban szociolingvisztikai háttérrel foglalkozott, míg Bourdieu inkább szociológiai szempontból vizsgálta a kérdést.

Bernstein szegényebb és gazdagabb gyerekek beszédhasználatát vizsgálta, nem a szókincs vagy a verbális képességek különbségei érdekelték, hanem a nyelvhasználat szisztematikus különbségei. Azt tapasztalta, hogy az alsóbb osztályok nyelvhasználatára a korlátozott kód jellemző, ami az jelenti, hogy a nyelvet sok olyan előfeltevéssel használják, amelyről azt felételezik, hogy a hallgató számára is ismertek, azaz olyan nyelvhasználatról van szó, ahol a mondanivaló nem választható le a helyzetről, amelyben létrejött. A felsőbb osztályok gyermekeire a kidolgozott kód használata jellemző, ami azt jelenti, hogy a mondanivaló leválasztható arról a helyzetről, amelyben létrejött, kevésbé kontextusfüggő, így ezek a tanulók könnyebben fejeznek ki általánosításokat és elvont fogalmakat is. Bernstein szerint azok a gyerekek, akik a kidolgozott kódot sajátították el, sikeresebben küzdenek meg az iskolai próbatételekkel, hiszen az oktatás kidolgozott kódban folyik, így ők egy ismerős nyelvi közeggel találkoznak, míg az alsóbb osztályok gyermekei könnyen kudarcként élik meg az iskolai nyelvhasználattal való találkozást. [8]

Bourdieu az iskola szerepéről írva egészen odáig megy, hogy az iskolának komoly része van a társadalmi szelekció fenntartásában – amelyben nyilvánvalóan szerepe van a Bernstein-féle nyelvhasználati különbségeknek is:

„A kiváltságos osztályok egyre teljesebb mértékben az iskolára ruházzák át szelekciós hatalmukat. Úgy tűnik, mintha ezzel egy teljesen semleges hatalom javára mondanának le a nemzedékek közötti hatalomátadás hatalmáról, s mintha feladnák a kiváltságok átörökítésének kiváltságát. Az iskola eljárása azonban a következő: formailag kifogástalan ítéleteket hoz. Ezek objektíve mindig az uralkodó osztályt szolgálják, hiszen még technikai érdekeit sem sértik soha – hacsak nem társadalmi érdekeik védelmében. Ilyen módon az iskola minden eddigénél jobban – s egy demokratikus ideológiára hivatkozó társadalomban az egyetlen elképzelhető módon működik közre a fennálló rend e reprodukciójában, mert minden eddigénél jobban leplezi el azt a funkciót, amelyet betölt...” [2]

Ha ilyen komoly különbségek lehetnek egy adott nyelvet használók között, akkor vajon alkalmazhatjuk-e ilyen esetekben a szóstatistikákról vázolt összefüggéseket, eredményeket fordítások összehasonlításánál? Ha igen, milyen megkötéseket kell tennünk az elmélet felhasználásával kapcsolatban, szükséges-e ilyen megkötések megtétele? Hogyan illeszthető be ez az eddig vázolt elméletbe? Ahhoz, hogy erre a kérdésre kielégítő választ kapjunk, magában a szövegben kell számba vennünk, hogy milyen tényezők befolyásolják a szóstatistikákban mutatkozó különbségeket. Ezek a tényezők a következők:

I. Eltérő nyelvi szerkezet. Az olasz nyelvet vizsgálva azt tapasztalhatjuk, hogy amikor egy bizonyos gondolatot megfogalmazunk, akkor a magyarhoz képest bizonyos szavaknak, szófajoknak az előfordulási gyakorisága minden esetben magasabb, ugyanannak a gondolatnak a magyar megfogalmazásához képest. Vannak azonban olyan szavak, amelyeknek a fordítást követően, azok magyarra fordításánál nem következik be lényeges változás. Vegyük sorjában a változásokat előidéző szerkezetbeli különbségeket! A felsorolás természetesen nem lehet teljes körű és részletekbe menő, hiszen az kimerítené ennek a tanulmánynak a kereteit, viszont fontos áttekintnünk azokat a főbb szerkezeti eltéréseket a két nyelv között, amelyek meghatározzák a szóstatistikában való elérést, hogy lássuk, hogy olyan szerkezeti sajátosságokról van

szó, amelyek szisztematikus eltérést mutatnak, és amelyek bizonyos keretek között kiszámíthatóvá teszi a magyar és az olasz szövegek statisztikai összehasonlítását.

Az igeik és az összetett igeidők használata során a szavak száma az olaszul megfogalmazott szövegrészekben gyakorlatilag szinte mindig magasabb lesz, mint azok magyar megfelelőiben. Ennek egyik oka az, hogy az olasz nyelv gyakran olyan esetekben is használja a létige ragozott alakját, amikor a magyarban ezt nem használjuk. Másrészt, míg a magyar nyelvben egyféle múlt időt használunk kijelző módban, az olaszban ötféleképpen fejezhetjük ki ugyanebben a módban egy bizonyos cselekvés múltidejét (*passato prossimo*-val, *imperfetto*-val, *trapassato prossimo*-val, *passato remoto*-val és *trapassato remoto*-val). Ezek közül az igeidők közül kettőnél az ige ragozott alakját használjuk (*imperfetto*, *passato remoto*), míg a másik három esetben összetett múlt időt. A *passato prossimo*-nál egy az *avere* vagy *essere* segédige jelen idejű ragozott alakjából és egy múlt idejű melléknévi igenévből álló szerkezettel fejezzük ki a múlt időt. Hasonló a helyzet a *trapassato prossimo*-nál és a *trapassato remoto*-nál is, azzal a különbséggel, hogy előbbinél az *avere* vagy *essere* segédigét *imperfetto*-ban, míg utóbbinál *passato remoto*-ban ragozzuk. A szenvedő és a műveltető szerkezet is az olasz szöveg szószámát növeli a magyar megfelelőhöz képest.

További eltérést eredményez a magyar és olasz szöveg szóstatisztikájában, hogy az olasz nyelv használ előljárószavakat, míg a magyar nyelvben ezek funkcióját a ragok és névutók tölti be. Amennyiben névelőt is használunk, az az olasz nyelvben összeolvad az előljárószóval. A részelő névelő használata is ismeretlen a magyar nyelvben.

A névmások használatában is jelölés elérés mutatkozik a magyar és az olasz nyelv között. Az olaszban ismeretlen a tárgyas igeragozás, így az olaszban a tárgyesetű személyes névmást minden esetben ki kell tennünk, amikor a magyarban tárgyas ragozást használnánk. A magyar nyelvben azonban a tárgyas ragozás már önmagában is kifejezi a tárgyat, így egyes szám 1. személyű alany és 2. személyű tárgy esetén nem szükséges kitenünk, mint ahogy az egyes szám 3. személyű tárgyat sem szükséges kitenni. A többes szám 2. és 3. személyű tárgyat viszont a magyarban is mindig kitesszük.

A birtokos jelző kifejezésére az olaszban olyan szerkezeteket használunk, amelyek szintén a szavak számának eltérését eredményezik a magyar nyelvű szerkezethez képest. Míg a magyarban a birtoklás kifejezésére a birtokhoz mindig birokos személyragot teszünk, az olaszban ezt kifejezhetjük egy birtok + *di* előljárószó + birokos szerkezettel, jelzői birtokos névmással (birokos determinánssal), valamint használhatunk személyes névmást a birtokos determináns helyett. „A birtokosra igen gyakran nem a birokos determináns, hanem a személyes vagy visszaható névmás *hangsúlytalan részes esete* utal. Ez jellegzetesen olaszos, a *magyartól teljesen eltérő* szerkezet.” [1]

Az olasz szövegben szereplő főnevek, melléknevek, számnevek általában a magyar szövegbe is ekként kerülnek lefordításra, de itt is lehetnek kivételek, például ha az olasz szövegben szereplő jelzős főnév a magyarban egy olyan kifejezést alkot, amely egy összetett szó, vagy ha az olasz főnév megfelelőjét a magyarban két szóban írjuk.

Az olasz és a magyar szövegrészek szószámbeli elérését leginkább a nyelvtani szerkezeti különbségeknek tulajdonítom, amelyek természetesen a gondolkozásbeli, szemléletmódbeli különbséggel függenek össze. Természetesen a továbbiakban leírt okok is közrejátszanak a szavak számának elérésében, de önmagukban nem lennének

elegendőek ahhoz, hogy egy jól behatárolt keretek között mozgó, szisztematikus eltérést vizsgáljunk.

II. Eltérő kifejezés- és gondolkodásmód. Amikor arról írtam, hogy a kétnyelvű japán nők gondolkodásmódja eltér, amikor japánul, illetve angolul kellett megválaszolniuk egy kérdést, eszembe jutott néhány olasz közmondás, amelyek magyarul egészen másképp hangzanának, ha szó szerint próbálnánk lefordítani őket, de néhány gyakori szófordulattal is ez a helyzet. A szófordulatok részben követik a nyelv eltérő szemléletmódjából adódó struktúrákat, részben pedig gyakoriságuk alapján vagy eleve szerepelnek a már meghatározott arányszámokban, vagy ha ritkán fordulnak elő, eleve nem befolyásolják lényegesen a kialakított eltérési arányszámokat.

III. Nyelvjárás, nyelvi rétegződés. A Wikipédia honlapján a következőket olvashatjuk az Olaszországban használt dialektusokról: „Az olasz dialektológia az újlatin nyelvészet egyik leggazdagabb területe. Olaszország területén számos (egyes becslések szerint 200 körüli) újlatin dialektust és aldialektust (nyelvjárást) használnak. Ezek az olasz dialektusok (*dialetti italiani*) északról dél felé haladva erősen különböznek egymástól, oly mértékben, hogy két távolabbi beszélő meg sem érti egymást: így a kölcsönös érthetőség végett mindenkinek beszélnie kell a sztenderd olasz nyelvet (olasz köznyelv).

Az erős dialektális tagolódás oka az egységes Olaszország, illetve az olasz irodalmi nyelv késői kialakulása volt. Sok olasz dialektust ma már teljesen önálló újlatin nyelvként tartanak számon, amelyek már saját helyesírással is rendelkeznek. Ilyenek a szicíliai, a nápolyi, az emilián-romanyol, a velencei, a lombard, a ligur, a piemonti, a szárd és a korzikai.” [9]

„Az olasz irodalmi nyelv alapjául a középolasz dialektuscsoport, ezen belül első-sorban a középkori toszkán dialektus szolgált. Az ebből kialakuló mai sztenderd olasz nyelvre azonban a többi közeli középolasz dialektus, így a római dialektus is hatást gyakorolt. Érdekes módon a mai toszkán dialektus a sztenderd olasz nyelvtől a kiejtésében észrevehetően elkülönül, például egész Olaszországban egyedül a toszkán dialektus használja a 'h'-mássalhangzót a 'k' helyett: például a 'come' szó *hóme* ejtése a köznyelvi *kóme* helyett. A köznyelvi 'cs'-mássalhangzót - a rómaihoz hasonlóan - a toszkán dialektus is 's'-nek ejti, szemben a sztenderd olasz ejtéssel: például a 'cinquecento' szó *sinkvesento* ejtése a köznyelvi *csinkvecsento* helyett.” [9]

Ahogy látjuk, egyrészt megkülönböztethetünk olyan dialektusokat, amelyek az idők folyamán önálló nyelvvé váltak, olyanokat, amelyek erősen eltérnek az irodalmi olasz nyelvtől, és olyat is, amely hangzóiban, egyes tájszávaiban tér el a mai olasz nyelvi sztenderdtől. Amikor egy dialektus annyira eltávolodik a nyelv általánosan bevett normáitól, hogy önálló nyelvvé válik, akkor erre a nyelvre külön meg kell határoznunk a szöveg fordításából adódó elérési arányt a szavakra. Ugyanez a helyzet az irodalmi, illetve a köznyelvtől való jelentős eltérés esetén is. Fordítási szempontból ezt az eltérést akkor tekinthetjük jelentősnek, ha a két nyelv szóstatisztikáinak eltérési rendje a meghatározott intervallumon kívülre esnek.

A nyelvi rétegződés szóstatisztikára gyakorolt hatásával foglalkozva szintén azt kell szem előtt tartanunk, hogy a nyelvi normáktól való eltávolodás milyen mértékben zajlik le, és mekkora hatással van az olasz és a magyar nyelv szóstatisztikája közötti különbségre. A korlátozott és a kidolgozott kód kérdéskörére visszaérve, azt feltételezem, hogy a korlátozott kódban elhangzó megnyilatkozások az adott nyelv nyelvtani szerkezetét követik, egy leegyszerűsített, hiányos szerkezetben, amely a magyar nyelv-

vű fordításnál valószínűleg olyan szisztematikus eléréseket eredményez, amely a megadott statisztikai határértékek közé esik. Megjegyzem ugyanakkor, hogy az írásban elhangzó közlések sajátossága, hogy kidolgozott kódban fogalmazódnak meg, és ha esetleg irodalmi környezetben meg is jelennek egy bizonyos társadalmi közeg bemutatására, hatásuk még ebben az esetben is elhanyagolható, hiszen egyrészt a szövegekörnyezet, amelyben szerepelnek kidolgozott kódban fogalmazódik meg, másrészt a cselekmény megértése szükségessé teszi, hogy a korlátozott kódot használók előfeltevéseit, a mondandót, amely nem választható le a helyzetről, amelyben létrejött, a szerző egyértelművé tegye az olvasó számára. Nem tartom azonban kizártnak, hogy a vizsgált szövegek között olyan jól körülhatárolható kategóriákat találjunk, amelyek statisztikai tulajdonságaikban eltérhetnek egymástól. Ezek azonban nem a felvázolt eljárás cáfolatai, hanem annak árnyalásai az eljárás szabályszerűségeinek felhasználásával.

III. Idegen nyelvű szövegek beékelődése a szövegbe. Amikor olasz nyelvű szöveget fordítunk magyarra, figyelembe kell vennünk a szövegben esetlegesen hosszabb terjedelemben szereplő idegen nyelvű idézeteket. Ha egy olasz szövegben, illetve annak részekre bontásánál egy szövegrészben hosszabb idegen nyelvű idézet van (például angol, német vagy francia), amit lábjegyzetben magyaráz meg a fordító, és ami így változatlan formában és szószámában kerül be a magyar szövegbe, a két szöveg statisztikai összehasonlítását nyilvánvalóan befolyásolhatja. Hosszabb idézet vagy beékelés esetén érdemes a vizsgált nyelvtől eltérő, idegen nyelvű szöveg nélkül összehasonlítani a szóstatisztikákat.

IV. A helyesírás aktuális szabályai. Amikor egy adott szöveget és annak fordítását vizsgáljuk, figyelembe kell vennünk az eltérési arányok meghatározásánál az egybe és különírás időszéri szabályait, amelyek időről időre változhatnak, valamint meg kell vizsgálnunk a szóban forgó nyelv és a magyar nyelv, adott korra jellemző nyelvhasználatát.

4 Összegzés

Olasz nyelvű szövegek magyar fordításának vizsgálata alapján igazolódni látszik az a feltevés, hogy az olasz szöveg szószáma szinte minden esetben magasabb a magyar szöveg szószámánál, és ez az eltérés jól behatárolható értékek között mozog. Ha bizonyos szövegrészek statisztikája a megadott határértékeken kívülre esik, akkor minden esetben szükséges az adott szövegész fordításának ellenőrzése. Amennyiben a megadott szövegrészek rendre eltérnek a megadott határértékektől, és a fordítás is helyes, akkor vagy olyan szöveggel van dolgunk, amelyre nem alkalmazható az eljárás (például képleteket, tudományos leírásokat, és anyagneveket felsoroló szöveg), vagy egy olyan szöveggel, amelyre eltérő statisztikai határértékek a mérvadók, amely az adott szövegtípusra is meghatározható (például nyelvjárási szöveg, rétegnyelv vagy egy jól körülhatárolható szövegtípus).

Feltételezhető, hogy az eljárás elsősorban a narratív struktúrákhoz kötött, az elbeszéléshez, és az elbeszélésben szereplő igeidők és szerkezetek váltakozásához, továbbá az is, hogy nemcsak a narratívum, hanem az azt tartalmazó szöveg formai sajátosságai is hordoznak egyfajta struktúrát, amely a narratívumot hordozó nyelv szerkeze-

tével és statisztikai jellemzőivel párhuzamosan vizsgálható, és felhasználható a más nyelvre fordított megnyilatkozás szerkezetével és statisztikai jellemzőivel való összevetésre. Mivel szisztematikus eltéréseket vizsgálunk, ezek más nyelvre is kiterjeszthetők, amelyre szintén külön meg kell határoznunk a szóstatisztikák közötti eltérés intervallumát.

A kutatás további lehetséges irányai: az eljárás alkalmazhatóságának vizsgálata más nyelveken és más szövegekre, valamint egy adott mű több különböző magyar nyelvű fordításának összevetése. Az így kapott eredmények tovább árnyalhatják a tanulmányban ismertetett módszert, és lehetőséget adhatnak az eredmények szélesebb körű felhasználására.

Hivatkozások

1. Angelini, M. T., Mórítz Gy.: Gyakorlati olasz nyelvtan, Nemzeti Tankönyvkiadó, Budapest, (2006)
2. Bourdieu, P.: A társadalmi egyenlőtlenségek újratermelődése. Tanulmányok. Fordította: Ádám P., Ferge Zs., Léderer P. Gondolt, Budapest, (1978)
3. Bruner, J.: A gondolkodás két formája. Forrás: László János, Thomka B. (szerkesztette): Narratívák 5. Narratív pszichológia. Kijárat Kiadó, Budapest, (2001) 27-58
4. Ciano, F.: Amikor nagyapa agyonlövete apát. Fordította: Puskás L. Kézirat.
5. Ciano, F.: Quando il nonno fece fucilare papà. A cura di Cimagalli, D.. Arnoldo Mondadori Editore, Milano, (1991)
6. Farb, P.: Word Play: What Happens When People Talk. Vintage Books, New York, (1993)(<http://cyberartsweb.org/cpace/theory/luco/Hypersign/Language.html>)
7. Forgács J.: A társas érintkezés szociálpszichológiája. Fordította: László J. Gondolat Könyvkiadó, Budapest, 2. kiadás, évszám nélkül.
8. Giddens, A.: Szociológia. Osiris Kiadó, Budapest. Fordította: Babarczy E., Nagy M., Nagy Zs., Tóth L. (1995)
9. Olasz nyelv. Forrás: http://hu.wikipedia.org/wiki/Olasz_nyelv